

Glossario

Marco Brugin

August 4, 2023

Abstract

Il documento contiene la descrizione di alcuni termini specifici del dominio di riferimento.

Glossary

Apache Hadoop è un framework open source che permette l'archiviazione e l'elaborazione distribuite di grandi moli di dati in cluster. Data la sua natura distribuita, presenta un'architettura estremamente scalabile e resiliente.

Base64 sistema di codifica che consente la traduzione di dati binari in stringhe di testo ASCII, rappresentando i dati sulla base di 64 caratteri ASCII diversi.

Consumatori (consumer) agenti che rivestono l'operazione di sottoscrizione a categorie di eventi e gli elaborano.

Container è un ambiente software completamente isolata che gira nativamente nel kernel della macchina ospitante. Un container contiene tutto il necessario per l'esecuzione dell'applicazione che ospita.

Data enrichment è il processo di arricchimento dei dati grezzi con informazioni aggiuntive, al fine di correggere piccoli errori, eliminare duplicati e migliorarne la qualità.

Data Injection metodo attraverso cui un insieme di dati vengono iniettati (inseriti) da un sistema software ad un altro.

Data warehouse è un sistema di archiviazione digitale che mette in collegamento e armonizza consistenti quantità di dati provenienti da molteplici fonti di vario tipo.

Docker Compose è uno strumento per definire e avviare applicazioni Docker multicontainer. È possibile utilizzare un file YAML per configurare i servizi dell'applicazione e, con un singolo comando, crearli ed avviarli da tale file. Il suo utilizzo è ristretto a esecuzioni su singola macchina.

Editori agenti che hanno il compito di estrarre il dato dall'evento e lo archiviano.

Event Handler un gestore di eventi è una routine di callback che opera in modo asincrono una volta che si verifica un evento.

Evento è la rappresentazione logica (segnali) che indicano che è successo qualcosa (un cambiamento di stato). Ogni evento è costituito da: una chiave, un valore, un timestamp e altri metadati facoltativi.

Hadoop è un framework open source che permette l'archiviazione e l'elaborazione distribuite di grandi set di dati in cluster di computer utilizzando semplici modelli di programmazione.

Kubernetes è un sistema di orchestrazione e gestione di container, sviluppato con il linguaggio di programmazione Go. Inizialmente sviluppato da Google (pubblicato nel settembre 2014), mantenuto ora da Cloud Native Computing Foundation. È in grado di funzionare con svariati sistemi di containerizzazione, tra cui Docker.

Minikube è un tool che permette di eseguire un cluster Kubernetes locale su una macchina virtuale.

OLAP acronimo di On-Line Analytical Processing, è un insieme di metodi finalizzato ad effettuare analisi rapide e approfondite su grandi volumi di dati. provenienti da uno o più data warehouse.

OLTP acronimo di Online transaction processing, indica un insieme di tecniche finalizzate a elaborare centinaia a migliaia di transazioni al secondo o più. Per creare la velocità e la precisione necessarie su larga scala, i processi OLTP utilizzano transazioni piccole e facili da eseguire che possono essere rapidamente verificate.

On-promise fornitura di programmi informatici installati e gestiti attraverso computer locali.

Origini indica l'origine del dato, da dove viene prodotto.

Postgres chiamato come PostgreSQL è un database relazione open source, molto utilizzato a livello mondiale, fornisce il supporto per diverse funzioni SQL, anche avanzate.

Produttori (producer) agenti che svolgono l'azione di pubblicare eventi su Kafka.

Sink agente che indicano la destinazione del dato.

Sottoscrittori agenti che hanno il compito di consumare e risponde all'evento generato con una azione in grado di gestirla .

Topic Topic o argomento non sono altro che cartelle del FS dove si archiviano gli eventi.